Using administrative databases in oncological population health research: a Canadian experience

D MacDonald¹,²*, R Alaghehbandan³, JC Knight¹,², J Rose⁴, KD Collins¹,²

Abstract

Introduction

In oncology research, there are important administrative data sources that could be used, different uses once the system is developed and generally opportunities and challenges associated with the use of administrative data for oncology research. Experiences gained through the development and implementation of a linked cancer surveillance database for studying cervical cancer are presented.

Methods

The database, referred to as the Cervical Cancer Surveillance System, contains linked cervical cancer case records, cervical cytology records, hospital discharge abstracts, physician claims and mortality records. The data linkage process was conducted in four separate but interconnected steps using the province unique health insurance number (MCP number) as the linking variable. The final database was a relational database with multiple interconnected tables that included patient, results and cervical cancer tables.

Discussion

There is limited literature on the challenges and opportunities of using administrative databases in oncological research; yet, the growing use of administrative data has increased the capacity of researchers to provide evidence to inform programmes and policies, including oncology care. However, linking data from various sources requires technical, ethical and privacy issues, and the time and resources needed to gain approval from data custodians and a research ethics board to access the data should not be underestimated. A multi-disciplinary research team with considerable experience in using administrative data developed an information system to study cervical cancer. The Cervical Cancer Surveillance System is a modular, multi-component system designed to accommodate different types of data from a variety of sources at the population level.

Conclusion

The Cervical Cancer Surveillance System supports the secondary use of administrative oncology data in the areas of surveillance, prevention, management, education, promotion and research.

Introduction

Administrative data are increasingly being used in applied health research, because they provide several advantages over primary data collection. Data collected for administrative purposes at a population level generally contain large numbers of records, include all events for a specific purpose and are more likely to be cost-effective relative to primary data collection.¹² Administrative databases have increasingly been used in studies on health service utilization, program evaluation and health outcomes, particularly in economically developed countries. Their use in health and social services research has enhanced the ability of administrators and policymakers to obtain information, which can be used to evaluate the impact of services and programmes on their populations. Administrative data as a valuable tool also serve to measure the impact of government programmes and services. The main disadvantage of administrative data is that the primary purpose for its collection is to support program or service delivery, not research. Governments need to consider collecting data beyond only that which is necessary to administer the program.

Cervical cancer is an almost entirely preventable cancer and one of the few cancers with effective screening modalities to identify cellular changes at the precursor level. Cervical cancer screening programmes require information systems to record, track and monitor a woman’s test and follow-up results and ultimately to measure the impact of screening programmes on cervical cancer incidence and mortality. Through a collaborative effort between a number of provincial organizations and institutes, a research team (including the authors) with diverse scientific and clinical backgrounds and expertise was formed, which ultimately developed an information system model for cervical cancer screening programmes. The paper presents the experiences gained through the development and implementation of the Cervical Cancer Surveillance System, a model that can be used and adapted to meet the needs of provincial-wide programmes across Canada.

Licensee OA Publishing London 2013. Creative Commons Attribution License (CC-BY)

**Materials and methods**

This section presents the framework and approach used to develop a provincial cervical cancer screening information system in Newfoundland and Labrador that can be used to monitor cervical cancer screening and outcomes, program evaluation and research.

This project was approved by the Human Investigation Committee at Memorial University of Newfoundland. The main goal of the project was to build a comprehensive longitudinal database for the study of cervical cancer and cervical screening in the province of Newfoundland and Labrador. The database, referred to as the *Cervical Cancer Surveillance System*, contains linked cervical cancer case records, cervical cytology records (Pap smears), hospital discharge abstracts, physician claims and mortality records. Linking cervical cancer records to these other administrative data sources allows researchers to study the relationship among screening rates, cervical cancer incidence and stage distribution, potential risk factors, related health care services utilization and health outcomes.

**Data sources**

The following administrative databases were used in the development of the *Cervical Cancer Surveillance System*: (1) Provincial Cancer Registry, hosted by the Provincial Cancer Care Program and includes demographic and clinical information on all cervical cancer patients in the province; (2) Provincial Cervical Cytology Registry, maintained by the Provincial Cancer Care Program and contains demographic data and cytological findings for Pap smears received from the provincial cytology laboratory; (3) Clinical Database Management System (provincial hospital separation data), maintained by the Centre for Health Information and includes demographic, clinical and procedural data collected by health agencies on all acute care and surgical day care hospitalizations in the province; (4) Mortality System, maintained at the Newfoundland and Labrador Centre for Health Information and includes information for a variety of conditions surrounding each death and (5) Medical Care Plan (MCP) fee-for-service physician claims data, maintained by the Government of Newfoundland and Labrador and includes claims records for fee-for-service physician visits in the province.

**Data linkage protocol**

The data linkage process was conducted in four separate but interconnected steps (Figure 1). In the first step, the cervical cancer and cytology data were imported into Microsoft Access 2000 software and linked using provincial patient unique health number (i.e. Medical Care Plan—MCP number) as the linking variable to form a relational database with multiple interconnected tables such as the “patient table” (patient’s demographic information), the “results table” (Pap smear findings) and the “cervical cancer table” (pathology diagnoses). In the second step, the provincial hospital data including data on each care episode (one record per hospitalization), diagnosis, procedure and

---

**Figure 1:** Model used for developing a comprehensive cervical screening information system.
Discussion

This article focuses on the feasibility and potential benefits of using administrative databases in oncology surveillance and research, using cervical cancer as an example. Effective and efficient delivery of cervical screening programs requires information for program planning, management, delivery, and evaluation. In most settings, special information systems are required to meet these needs. Ideally, organized screening programs should have information systems that cover all aspects of screening and follow-up. Regardless, without an effective information system, it is not possible to run an organized cancer screening program and evaluate whether it is operating as expected, achieving program objectives, or supporting effective research. Information systems are often considered stand-alone products when they are designed and implemented, but to achieve maximum benefit they should be integrated into existing health care information systems as a means to provide comprehensive data for patient care, research, and program evaluation.

The growing use of administrative data in applied health research in economically developed countries has increased the capacity of researchers to provide evidence of the impact on population health of various health programs and policies, including oncology care. Administrative databases are a rich source of data that can be used to investigate what services were utilized, how these services were used and the broader impact of these services on society. However, answering the questions that researchers and/or policymakers pose often requires that data from two or more databases be combined, given that the data required are often not available from a single source.4

Administrative databases typically contain large numbers of records with detailed demographic and service event data, lending themselves to sophisticated statistical analyses. They are considered a relatively low-cost source of data compared with primary data collection and are often longitudinal by design, which supports trending (or Panel) analysis. Administrative data can be even more useful when two or more sources are linked together; the result of which is a more comprehensive compilation of data. However, linking data from various sources raises technical issues such as the compatibility of operating systems and the availability and quality of common identifiers in the databases to be linked.

Administrative databases have been used in oncology research studies including breast, gastrointestinal and head/neck cancers.5-7 However, there is limited knowledge in the literature with regard to challenges and opportunities of using administrative databases in oncological population health research. Our experience highlights a feasible methodology and framework in developing such a surveillance system for cervical screening and cancer.

With respect to oncological population surveillance, the following goals for information system requirements, components and functionality are essential, including populations at risk, maintaining current information, providing follow-up, supporting quality assurance, tracking utilization and monitoring compliance with screening recommendations. For a cervical cancer prevention program, an information system requires the following data in order to serve as a program management tool:

- Information on the target population, including the number of women in the target age group
- Screening test results
- Data from follow-up, management and outcomes of positive test results
- Cervical cancer diagnosis and deaths
- Service utilization

The systems should contain information from various data sources to assist in program management and to generate performance indicators. In this project, different data sources (e.g., cancer, cytology, hospital, mortality, etc.) were used. Other potential data such as national health surveys and laboratory findings on human papilloma virus status can be linked to the existing cervical screening database. Further, the database could be linked to health insurance registry/client registry to allow inclusion of individuals without cancer; as well as those who have not been screened, to allow examination of risk factors for cervical cancer and screening.

An important consideration when developing a population health database is the need to protect individuals’ privacy. It should be noted that as more disparate databases are linked, the higher the potential for identification of an individual in the resulting linked database. Most administrative databases contain data that, by law, have restrictions on their permitted uses. Researchers who seek access to these data must negotiate with the custodian of the data, obtain approval of a research ethics board (REB) and demonstrate how they will meet the legal criteria for access. Ethics criteria generally require that the investigators identify the benefits associated with the release of the data, and in particular, how the data will be protected from unauthorized use.
access and disclosure. The time and resources needed to gain approval from data custodians and an REB should not be underestimated. The researcher may be required to work with multiple data custodians to gain approval for access, and each custodian may have different policies and processes in place related to access, which adds to the complexity and timeliness of the study. It is also possible that a data custodian will charge a fee for accessing the data, and the researcher should account for this cost in their research budget. In this initiative, the research team included pertinent stakeholders and parties from various organizations and institutions, which facilitated the process. Further, it is clear that when conducting applied health research using administrative data, it is important to have people on the research team that have knowledge of both the database(s) being used and the various limitations that present when using administrative data. In this project, a multidisciplinary team of researchers, including those experienced in the use of administrative databases, as well as researchers having clinical, social and economic expertise, was assembled.

The challenges continue once the researcher gains access to the data, as issues with data quality, such as missing, incomplete or inaccurate data, need to be considered. Quality assurance processes associated with administrative databases are not as stringent as those developed primarily for research. It is important to recognize that even when data quality issues have been identified, it is unlikely they can be corrected, given that the researcher will not have access to the primary data collection records.

One of the other challenges that arise when using administrative data for research is that most of these databases are primarily developed to deliver public programmes or services. The authors believe that there is a need to encourage governments at all levels to consider potential secondary uses of administrative data when developing any initiative that involves the creation of new databases/information systems.

Conclusion

The information system model created in this project was a modular, multi-component system designed to accommodate different types of data from a variety of sources at the population level. It should be noted that while building the model, desirable system characteristics were identified; these guided subsequent iterations to the model and implementation strategies including functionality, simplicity, sustainability, flexibility and data availability.

Acknowledgements

This project was financially supported by the Newfoundland Cancer Treatment Research Foundation (NCTRF), Eastern Regional Health Authority and the Cervical Screening Initiatives (CSI) program, Western Regional Health Authority, NL, Canada. The authors thank Terry Malone at the Provincial Cancer Care Program for his assistance throughout the project.

Abbreviations list

MCP, Medical Care Plan; REB, research ethics board.

References